AERA 2018 Symposium Proposal

Teaching Critical Thinking: Assessing and Improving Students' and Teachers' Reasoning Skills

Symposium Abstract

Fostering students' critical thinking (CT) skills is one of the major ambitions of education. However, there are many open questions with regard to teaching CT, such as how to assess students' reasoning skills, what effective instructional interventions would be (especially for establishing transfer) and whether teachers are and feel equipped to teach CT. The first two contributions in this symposium focus on assessing students' reasoning skills with a domain-specific test and in vivo during team problem solving. The last two contributions present experimental studies on the effectiveness of instructional interventions to improve students' and teachers' reasoning skills. Our discussant, Karen Murphy, will engage the authors and audience in critical thinking about these studies.

Symposium Summary

Fostering students' critical thinking (CT) skills is one of the major ambitions of education (National Research Council, 2012). This is not surprising, as CT-skills have been shown to positively relate to learning outcomes (Arum, Cho, Kim, & Roksa, 2011), and because a lack of CT-skills can result in erroneous decisions that may have serious consequences, both in daily life and in highly complex professional domains (e.g., Lunn, 2011; Schmidt et al., 2016; Toplak, Liu, Macpherson, Toneatto, & Stanovich, 2007). What is surprising, is that CT is seldom explicitly taught; it is often expected to materialize as a by-product of education (Jones, 2007). However, to realize this ambition of delivering students who are critical thinkers, explicit teaching is necessary, as research has shown that CT-skills do not develop spontaneously from immersion in (higher) education (Abrami et al., 2015; Arum & Roksa, 2011; Pascarella, Blaich, Martin, & Hanson, 2011). However, there are many open questions with regard to teaching CT, such as how to assess students' reasoning skills, what effective instructional interventions would be (especially for establishing transfer), and whether teachers are and feel equipped to teach CT. In the four contributions to this symposium, we present new empirical research on those questions.

The first two contributions focus on assessing students' reasoning skills. Sermeus and Elen present data on a test they developed to assess secondary education students' domain-specific CT-skills in physics. Jablansky, Alexander, and Schmidt investigate manifestations of higher education students' relational reasoning during team problem solving, with the aim of developing instructional interventions.

The last two contributions present experimental studies on the effectiveness of instructional interventions to improve higher education students' and teachers' CT-skills, more specifically the ability to avoid biases in reasoning. Van Peppen, Verkoeijen, Heijltjes, Janssen, Kolenbrander, and Van Gog conducted two experiments to investigate whether interleaved (as compared to blocked) practice schedules with worked examples (as compared to practice problems) would foster students' learning and transfer of rational reasoning skills. Janssen, Mainhard, Heijltjes, Verkoeijen, Van Peppen, and Van Gog present a study on the effects of a training on teachers' learning and transfer of reasoning skills, their ability to recognize biases in student products (essays), and their attitude towards teaching CT.

Finally, we are pleased that Karen Murphy agreed to act as our discussant. She will engage the authors and audience in critical thinking about these studies.

Chair: Tamara van Gog
Discussant: Karen Murphy

Contributions:

1. From Domain-General to Domain-Specific Assessment of Critical Thinking: The Case of Electricity And Magnetism

Jan Sermeus[1], Mieke de Cock[1], and Jan Elen[1]

[1]University of Leuven

Paper Abstract
Introduction:
Critical thinking (CT) is one of the most desirable objectives of most, if not all, secondary and tertiary education curricula. The question then arises: how can CT be assessed? In this work we try to, in part, address this question. CT-tests are, at this moment, predominantly domain-general. However, some authors stress the importance of domain-specific knowledge for CT or stress the domain-specificity of CT (McPeck, 1981; Moore, 2004). Hence, there is a need for domain-specific tests to advance the field.

Confronted with educational goals that focus on CT on the one hand and the lack of validated tests to test domain-specific CT on the other hand, this study aims at constructing and validating a domain-specific test for physics (more specifically for electricity and magnetism – E&M). The test is aimed at secondary school students that have received instruction on E&M.

Method:
The development of the test followed design guidelines (Adams, & Wieman, 2010) in which several distinct phases in the development can be distinguished:
1) Construction of the theoretical framework based on Halpern's construction of CT.
2) Construction of the test-items reflecting the structure of the Halpern critical thinking assessment, HCTA (Halpern, 2014).
3) Focus-groups discussion with in-service teachers regarding the understandability and answerability of the test-items for students of the targeted population.
4) Cognitive interviews (Willis, 2004) with students of the targeted population.
5) Large group administration (convenience sampling, N=162).

Results:
Internal consistency of the test was found to be $\lambda_6 = 0.639$. This value is relatively low, a value of 0.7 is often set as minimum (Sijstsma, 2009). However, for a test that measures complex cognitive abilities, as certainly CT is, a lower internal consistency might not only be expected but also wanted (Peters, 2014).

Given the time constraints only two thirds of the students (N=105) were able to complete the entire test. On average the students scored 9 ± 4.5 out of 46 points.

Discussion and Conclusion:
Even though the sample in this study cannot be said to be representative for all students in the targeted population, the low average score obtained by the participating students raises multiple questions:

1) Was the test was too difficult for the students? This seems unlikely as in stage 3 of the test-development the in-service teachers confirmed that all questions should be feasible for their students.
2) Are students not adequately prepared to think critically in their school career? A positive answer would be quite worrisome as this means that a major educational goal was not reached.
3) Is the test not valid as it is based on the translation of a domain-general framework to a specific domain?

Future research should try to answer the questions asked above, for example by comparing the domain specific CT skills probed in this test with domain general CT skills probed in the HCTA. Future research is also needed to design an educational methodology to increase the CT skills and disposition among secondary education students. An infusion approach seems promising (Davies, 2006; Tiruneh et al., 2016).

2. Relational Reasoning in Engineering Design Teams: Establishing a Baseline for

Sophie Jablansky[1], Patricia Alexander[1], and Linda Schmidt[1]

[1]University of Maryland, College Park

Paper Abstract
Objectives and Theoretical Framework:
Teams, rather than individuals, are commonly called upon to solve difficult problems in the workplace (Hackman, 1992). One area in which teams are core to the discipline and to academic programs is engineering design, which focuses on solving complex and ill-structured problems (DeChurch & Mesner-Magnus, 2010). Although design teams are commonplace in engineering, little is known about the role relational reasoning plays in performance or how instructional programs might be better devised to enhance such reasoning. The limited research that does exist suggests that design teams exhibit various forms of relational reasoning, including analogical (similarity), anomalous (deviation), antithetical (opposition), and antinomous (incompatibility) reasoning (Christensen & Ball, 2016; Dumas et al. 2014). However, there is limited understanding about how forms of relational reasoning play out as teams actively engage in problem solving or how the effectiveness of teams reflects their reasoning patterns. Without such a critical baseline, it is not evident how instructional programs involving design teams can be reconfigured to ensure their effectiveness. For that reason, the current investigation undertook variable- and person-centered analyses to explore the occurrence of relational reasoning within and between teams and to characterize teams on the basis of a cluster of variables related to facets of reasoning patterns, team discourse, and design quality indicators.

Method and Data Sources:
Participants were three teams (n=17) of undergraduates enrolled in a senior engineering design course. Teams were videotaped in their first three meetings, which took place outside of class. In these meetings, teams were tasked with conceptualizing ten viable designs, then narrowing

those down to three, and then to one. Content analysis was performed on student discourse by coding relational reasoning utterances using a framework developed by Dumas et al. (2014). In addition to reasoning data, information collected included the quality of ideas (i.e., innovativeness, viability) forwarded by each team, two graded reports, and final course grades.

Results:
For the variable-centered analyses, chi-square tests of independence revealed that all four forms of relational reasoning were present but were disproportionally verbalized across meetings and teams. Further, a transition matrix demonstrated that certain forms of relational reasoning transitioned to other forms with a higher frequency than expected by chance. To explore the unfolding of relational reasoning in more depth, a sequence mining tool was constructed. The tool revealed that analogies sparked more analogies in strings of up to six, and antinomies lead to strings of analogies or antitheses. For the person-centered analysis, we characterized the three teams using the available data as highly efficient, moderately effective, and disparate reasoners/designers. The presentation concludes with implications for course development.

Scientific Significance:
This *in vivo* study is the first, to our knowledge, to examine manifestations of relational reasoning forms during team problem solving. Findings revealed that team effectiveness in engineering design was supported by the usage of different forms of reasoning in concert that helped in task completion. Importantly, these findings serve as a baseline upon which future curricular enhancements might be considered.

3. Learning to Avoid Biased Reasoning: Effects of Interleaved Practice and Worked Examples

Lara van Peppen[1], Stefan Kolenbrander[2], Peter Verkoeijen[1,2], Anita Heijltjes[2], Eva Janssen[3], and Tamara van Gog[3]

[1]Department of Psychology, Education and Child Studies, Erasmus University Rotterdam
[2]Learning and Innovation Center, Avans University of Applied Sciences, Breda
[3]Department of Education, Utrecht University

Paper Abstract
Fostering critical thinking (CT) skills is an important educational objective. CT-skills are pivotal for learning, but also for daily-life decision-making and complex professional skills. One important aspect of CT is avoiding biased reasoning. Recent research showed that explicit instructions combined with practice were effective for learning to avoid biased reasoning, but did not foster transfer to non-practiced tasks (Heijltjes et al., 2015). We therefore investigated whether transfer improves from interleaved practice with worked examples.

According to the *contextual interference effect*, transfer is enhanced when materials are presented and learned under conditions of high interference (Schneider et al., 2002). Contextual interference can be created by interleaved as opposed to blocked practice. Whereas blocked practice involves practicing one task-category at a time before the next, interleaved practice mixes practice of several categories. Interleaved practice allows students to infer the distinctive characteristics of different problem-types and to select a learned procedure for solving new problems (Kornell & Bjork, 2008). However, interleaved practice is usually more cognitively

demanding than blocked practice. Hence, when tasks impose lower cognitive load –such as worked examples that present the worked-out solution procedure– more cognitive capacity is available for comparing problem categories (Paas & Van Merriënboer, 1994). Consequently, students might benefit more from high contextual interference when practicing with worked examples than with practice problems.

In two experiments, we investigated the effects of practice schedule (blocked vs. interleaved) and practice-task format (worked examples vs. problems) on learning and transfer of unbiased reasoning. 142 (experiment 1) and 109 (experiment 2) higher education students first learned how to solve base-rate (Br), conjunction (C), and syllogistic reasoning (S) tasks by video instruction followed by practicing 9 tasks. Depending on assigned condition, they practiced in a: 1) blocked schedule (e.g. Br-Br-Br-C-C-C-S-S-S) with worked examples; 2) interleaved schedule (e.g. Br-C-S-C-S-Br-S-Br-C) with worked examples; 3) blocked schedule with problems; or 4) interleaved schedule with problems. Thereafter, participants completed an immediate and delayed (two weeks later) posttest on five task-categories, to measure effects on learning (i.e. performance on practiced task-categories) and transfer (i.e. performance on non-practiced tasks-categories: contingency and Wason selection).

Participants learned, i.e., they improved from pretest to immediate and delayed posttest on practiced tasks, but there was no effect of practice schedule. There was an effect of practice-task format, but it was inconsistent across experiments: worked examples were more beneficial in experiment 1, and practice problems in experiment 2. The predicted three-way interaction between test-moment, practice schedule and practice-task type did not emerge. Regarding *transfer* to non-practiced tasks, experiment 1 showed improvement from immediate to delayed posttest when practicing with problems in an interleaved schedule. However, experiment 2 failed to replicate this interaction effect, showing no effects on transfer.

Thus, creating high interference during practice did not enhance learning and transfer of unbiased reasoning in our study, possibly because the task-categories were highly distinctive. Effects of worked examples were inconsistent across student populations. Future research should continue to search for conditions under which transfer would be enhanced.

4. Teacher Training on Critical Thinking: Effects on Rational Reasoning and Teaching Attitudes

Eva M. Janssen[1], Tim M. Mainhard[1], Anita E. G. Heijltjes[2], Peter Verkoeijen[2,3], Lara M. van Peppen[3], & Tamara van Gog[1]

[1]Department of Education, Utrecht University
[2]Learning and Innovation Center, Avans University of Applied Sciences, Breda
[3]Department of Psychology, Education and Child Studies, Erasmus University of Rotterdam

Paper Abstract
Although every review on how to improve students' Critical Thinking (CT) highlights the crucial role of the teacher in this process (e.g., Abrami et al., 2015; Ritchhart & Perkins, 2004), research on teachers' CT-skills is scarce and mostly focused on pre-service teachers. A prerequisite for being able to provide instruction and guidance to students on a subject is that teachers themselves possess the required skill but also that they perceive it as a highly relevant to teach it and that they identify themselves as self-competent in teaching it (Posnanski, 2002; Van Aalderen-Smeets & Walma Van der Molen, 2015). Therefore, the present experiment

investigated the effects of a teacher training in CT, focusing on biases in reasoning, on higher education teachers' performance on a rational reasoning test and their attitudes towards teaching CT (distinguishing between perceived relevance of and perceived competence in teaching CT).

Participants were 56 in-service teachers from a Dutch University of Applied Sciences: 34 teachers participated in a CT-course of three sessions spread over six weeks and 22 did not receive any training (i.e., control condition). The first course session took about 120 minutes and consisted of a general introduction of CT and explicit instruction on cognitive biases in logical and probabilistic reasoning. The remaining two sessions took approximately 180 minutes and 120 minutes, respectively, and focused on the teaching of CT (e.g., designing a domain-specific CT-task and discussion about what type of questions evoke students' CT). All teachers completed pre-test measures via an online questionnaire two weeks before the start of the first training session. Teachers in the training conditions completed post-test measures on two occasions: immediately after the first session and after the third session. Teachers in the control condition received the two post-test measures via an online questionnaire during the same weeks as the teachers in the training condition. All tests included a rational reasoning test with learning and transfer tasks; an assignment to detect cognitive biases in a student product (essay); and a questionnaire that addressed teachers' attitudes towards teaching CT (perceived relevance and perceived competence).

Results showed that the CT-training positively affected teachers' rational reasoning skills on trained tasks, $F(2,96)=5.84$, $p=.004$, $\eta_p^2=0.108$. Transfer to other (non-trained) logical and probabilistic reasoning tasks was not achieved, $F(2,96)=0.32$, $p=.724$, $\eta_p^2=0.007$. Data on whether the training affected teachers' ability to recognize biases in student products are currently being analyzed. Regarding attitudes, the training did not affect teachers' relevance perception of teaching CT, $F(2,94)=2.36$, $p=.100$, $\eta_p^2=0.048$, presumably because of a ceiling effect as all teachers consistently perceived teaching CT highly as relevant. Teachers' competence perceptions temporarily dropped after the first training session, $t(33)=5.11$, $p<.001$, possibly because this first session increased their awareness of how challenging CT actually is, which may have led to uncertainty about their ability to teach it.

Our findings underline that we cannot assume that teachers are equipped for and feel competent in teaching CT. Future research should therefore continue to focus on how to optimally prepare and support teachers in teaching CT-skills to students.

References

Abrami, P. C., Bernard, R. M., Borokhovski, E., Waddington, D. I., Wade, C. A., & Persson, T. (2015). Strategies for teaching students to think critically: A meta-analysis. *Review of Educational Research*, *85*, 275-314. doi:10.3102/0034654314551063

Adams, W. K., & Wieman, C. E. (2011). Development and validation of instruments to measure learning of expert-like thinking. *International Journal of Science Education*, *33*(9), 1289-1312. doi:10.1080/09500693.2010.512369

Arum, R., & Roksa, J. (2011). *Academically adrift: Limited learning on college campuses*. University of Chicago Press.

Christensen, B. T. & Ball, L. J. (2016). Creative analogy use in a heterogeneous design team: The pervasive role of background domain knowledge. *Design Studies, 46,* 38-58.

Davies, W. M. (2006). An 'infusion' approach to critical thinking: Moore on the critical thinking debate. *Higher Education Research & Development*, *25*(2), 179-193. doi: 10.1080/07294360600610420

DeChurch, L. A. & Mesner-Magnus, J. R. (2010). The cognitive underpinnings of effective teamwork: A meta-analysis. *Journal of Applied Psychology, 95,* 32-53.

Dumas, D., Alexander, P. A., Baker, L. M., Jablansky, S., & Dunbar, K. N. (2014). Relational reasoning in medical education: Patterns in discourse and diagnosis. *Journal of Educational Psychology, 106,* 1021-1035. http://dx.doi.org/10.1037/a0036777

Hackman, J. R. (1992). Group influences on individuals in organizations. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (pp. 199-267). Palo Alto, CA, US: Consulting Psychologists Press.

Halpern, D. F. (2010). Halpern critical thinking assessment. *SCHUHFRIED (Vienna Test System): Moedling, Austria. Http://www. Schuhfried. com/vienna-Test-System-vts/all-Tests-fromaz/test/hcta-Halpern-Critical-Thinking-Assessment*, *1*.

Heijltjes, A., Van Gog, T., Leppink, J., & Paas, F. (2015). Unraveling the effects of critical thinking instructions, practice, and self-explanation on students' reasoning performance. *Instructional Science*, *43*, 487–506. doi:10.1007/s11251-015-9347-8

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories is spacing the "enemy of induction"?. *Psychological science*, 19, 585-592. doi: 10.1111/j.1467-9280.2008.02127.x

Lunn, P. D. (2013). The role of decision-making biases in Ireland's banking crisis. *Irish Political Studies*, *28*, 563-590. doi:10.1080/07907184.2012.742068

McPeck, J. E. (1981). *Critical thinking and education.*

Moore, T. (2004). The critical thinking debate: how general are general thinking skills?. *Higher Education Research & Development*, *23*(1), 3-18. doi: 10.1080/0729436032000168469

National Research Council (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st century.* Committee on Defining Deeper Learning and 21st Century Skills, J.W. Pellegrino and M.L. Hilton, Editors. Board on Testing and Assessment and Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Paas, F. G., & Van Merriënboer, J. J. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of educational psychology*, *86*, 122.doi.org/10.1037/0022-0663.86.1.122

Pascarella, E. T., Blaich, C., Martin, G. L., & Hanson, J. M. (2011). How robust are the findings of academically adrift? *Change: The Magazine of Higher Learning*, *43*, 20-24.

Peters, G. J. Y. (2014). The alpha and the omega of scale reliability and validity: why and how to abandon Cronbach's alpha and the route towards more comprehensive assessment of scale quality. *European Health Psychologist*, *16*(2), 56-69. Retrieved from http://www.ehps.net/

Pithers, R. T., & Soden, R. (2000). Critical thinking in education: A review. *Educational Research*, *42*, 237-249. doi:10.1080/001318800440579

Posnanski, T. J. (2002). Professional development programs for elementary science teachers: An analysis of teacher self-efficacy beliefs and a professional development model. *Journal of Science Teacher Education*, *13*, 189-220. doi:10.1023/A:1016517100186

Ritchhart, R., & Perkins, D. N. (2005). Learning to think: The challenges of teaching thinking. *The Cambridge handbook of thinking and reasoning*, 775-802.

Schmidt, H. G., van Gog, T., Schuit, S. C., Van den Berge, K., Van Daele, P. L., Bueving, H., ... & Mamede, S. (2016). Do patients' disruptive behaviours influence the accuracy of a doctor's diagnosis? A randomised experiment. *BMJ Quality & Safety*, *0,*1-5. doi:10. 1136/bmjqs-2015-004109

Schneider, V. I., Healy, A. F., & Bourne, L. E. (2002). What is learned under difficult conditions is hard to forget: Contextual interference effects in foreign vocabulary acquisition, retention, and transfer. *Journal of Memory and Language*, 46, 419-440. doi.org/10.1006/jmla.2001.2813

Tiruneh, D., Weldeslassie, A., Kassa, A., Tefera, Z., De Cock, M., Elen, J. (2016). Systematic design of a learning environment for domain-specific and domain-general critical thinking skills. *Educational Technology Research and Development, 64* (3), 481-505. doi:10.1007/s11423-015-9417-2

Toplak, M. E., Liu, E., MacPherson, R., Toneatto, T., & Stanovich, K. E. (2007). The reasoning skills and thinking dispositions of problem gamblers: A dual-process taxonomy. *Journal of Behavioral Decision Making*, *20*, 103-124. doi:10.1002/bdm.544

Van Aalderen-Smeets, S. I., & Walma van der Molen, J. H. (2015). Improving primary teachers' attitudes toward science by attitude-focused professional development. *Journal of Research in Science Teaching*, *52*, 710-734. doi:10.1002/tea.21218

Willis, G. B. (2004). *Cognitive interviewing: A tool for improving questionnaire design*. Sage Publications